Novel Polypeptides and Methods for their Use

Government Funding

The invention described herein was made with U.S. Government support under United States Public Health Service grant number R01 GM29909 awarded by the National Institute of General Medical Sciences. The United States Government has certain rights in the invention.

Background of the Invention

Interest in the substrate specificity of bacterial dioxygenases stems from initial studies on the degradation of benzene and toluene more than 25 years ago. A mutant strain of *Pseudomonas putida* (strain F39/D) was shown to oxidize benzene and toluene to *cis*-1,2-dihydroxycyclohexa-3,5-diene (*cis*-benzene dihydrodiol) and *cis*-(1*S*, 2*R*)-dihydroxy-3-methylcyclohexa-3,5-diene (*cis*-toluene dihydrodiol), respectively (D.T. Gibson, et al., *Biochemistry*, **1970**, *9*, 1631-1635; D.T. Gibson, et al., *Biochemistry*, **1970**, *9*, 1626-1630; and V.M. Kobal et al., *J. Am. Chem. Soc.*, **1973**, *95*, 4420-4421).

The enzyme catalyzing these reactions, toluene dioxygenase (TDO), is capable of producing enantiomerically pure cyclohexadiene *cis*-diols from a wide range of aromatic substrates. D.T. Gibson, et al., *Microbial Degradation of Organic Compounds* (Gibson, D. T., ed.), pp. 181-251, Marcel Dekker, New York, NY (1984); D.T. Gibson, et al., *Pseudomonas: biotransformations, pathogenesis, and evolving biotechnology*, (Silver, S. et al., ed.), pp. 121-132, American Society for Microbiology, Washington D. C. (1990); G.N. Sheldrake, *Chirality in Industry: the Commercial Manufacture and Application of Optically Active Compounds* (Collins, A. N. et al., eds.), pp. 127-166, John Wiley & Sons, Chichester, UK (1992); Stabile, M. R., Ph.D. thesis. Virginia Polytechnic Institute and State University, Blacksburg, VA (1995); and D.R. Boyde and G.N. Sheldrake, *Nat. Prod. Rep.* **1988**, *15*, 309-324.

In contrast to the body of work relating to TDO, relatively little attention has been paid to the related enzyme naphthalene dioxygenase. Cells of *Pseudomonas* sp NCIB 9816-4 contain an inducible multi-component enzyme system designated NDO, which initiates naphthalene catabolism by catalyzing the addition of both atoms of molecular oxygen and two hydrogen atoms to the substrate to yield enantiomerically pure (+)-*cis*-(1*R*,2*S*)-dihydroxy-1,2-dihydronaphthalene (D.M. Jerina et al. *Arch. Biochem. Biophys.* **1971**, *142*, 394-396). NDO has a relaxed substrate specificity and catalyzes the deoxygenation of many related 2- and 3-ring aromatic and hydroaromatic (benzocyclic) compounds to their respective *cis*-diols.

The potential of NDO to form products of opposite chirality to those formed by TDO was first noted in 1988 during studies on the oxidation of indan. The major product formed by TDO was (−)-(1*R*)-indanol (84% enantiomeric excess [ee]) whereas NDO produced (+)-(1*S*)-indanol (>92% ee) (L.P. Wackett et al., *Biochemistry*, **1988**, *27*, 1360-1367. Subsequent studies with NDO revealed further differences in substrate specificity and suggested that this enzyme is an additional source of chiral intermediates and synthons for the enantiospecific synthesis of biologically active products. S.M. Resnick et al. *Journal of Industrial Microbiology*, **1996**, *17*, 438-457.

NDO belongs to a family of bacterial enzymes that have an essential role in the recycling of carbon in nature. These enzymes are especially important in the degradation of aromatic hydrocarbons and related environmental pollutants. Knowledge of the NDO reaction mechanism is thus important in the development of bioremediation strategies for cleaning up environments contaminated with hazardous aromatic compounds. An attractive alternative to bioremediation is the application of 'green chemistry,' which refers to the production of industrial chemicals by processes that do not generate hazardous waste. For example, a recombinant strain of *Escherichia coli* expressing NDO, has been used to synthesize indigo dye from glucose. *cis*-Arene diols produced by NDO and toluene dioxygenase have been used in the synthesis of many products of biological and economic importance.

Knowledge of the types of reactions catalyzed by NDO and the range of substrates oxidized by NDO is based largely on biotransformation studies with *cis*-naphthalene dihydrodiol dehydrogenase (DDH) mutants, recombinant strains expressing NDO and purified NDO components. *Pseudomonas* sp 9816/11 is a DDH mutant of strain 9816-4 (G.M. Klecka and G.T. Gibson, *Biochem J.*, **1979**, *180*, 639-645) which accumulates *cis*-naphthalene-1,2-dihydrodiol when induced cells are incubated with naphthalene and a suitable carbon source (D.S. Torok, *J. Bacteriol.* **1995**, *177*, 5799-5805. Studies with purified dioxygenase components have been crucial in the identification of reactions catalyzed by NDO in the absence of other host-associated enzyme activities which, through subsequent catalysis, have the potential to affect product distribution and/or stereochemistry.

In addition to *cis*-dihydroxylation, NDO also catalyzes a variety of other oxidations which include monohydroxylation, desaturation (dehydrogenation), *O*- and *N*-dealkylation and sulfoxidation reactions. S.M. Resnick et al. *Journal of Industrial Microbiology*, **1996**, *17*, 438-457. Many of the reactions catalyzed by NDO and other microbial dioxygenases yield hydroxylated compounds that can serve as chiral intermediates or chiral synthons for a variety of compounds of interest to pharmaceutical and specialty chemical industries.

Despite the wide range of useful oxygenated materials that can be prepared with TDO and NDO, there is currently a need for additional oxygenated chiral synthons that can be used to prepare therapeutically useful compounds, or useful intermediates. In particular, there is a need for additional chiral synthons that differ from the TDO and NDO products by absolute configuration or by the site of oxygenation. There is also a need for new methods to prepare hydroxylated aryl compounds for use in the polymer, resin, pharmaceutical or rubber industry, which generate less industrial waste than currently available methods. Further, there is a need for novel enzymes possessing structures, stabilities, or reactivities that differ from the native enzymes.

Summary of the Invention

The crystal structure of NDO has recently been published by B. Kauppi et al. *Structure*, **1998**, *6*, no. 5, 571-586. Based on this structure, the amino acid at position 352 is located at the active site of NDO. As described hereinbelow, site-directed mutagenesis was used to construct DNA molecules that encode NDO mutants having amino acid substitutions at position 352. Changing the amino acid at position 352 from phenylalanine to valine provided an enzyme (SEQ ID NO:2, encoded by SEQ ID NO:1) that gives a change in the preferred absolute configuration of the 1,2-dihydroxy-1,2-dihydronaphthalene formed from naphthalene. This enzyme also gave a change in the regioselectivity of the products obtained from oxidation of biphenyl and phenanthrene.

Accordingly, the invention provides an NDO or NDO related complex comprising a plurality of polypeptides, wherein the complex comprises at least one alpha-subunit polypeptide that comprises: 1) a substituted amino acid (e.g. valine or leucine) at the position corresponding to position 352 in NDO, 2) a substituted amino acid at the position corresponding to position 201, 202, 260, 316, 351, 358, 362, or 366 in NDO, or 3) a substituted amino acid at the position corresponding to position 352 in NDO, and a substituted amino acid at the position corresponding to position 201, 202, 260, 316, 351, 358, 362, or 366 in NDO; or a catalytically active fragment thereof. The complexs of the invention can preferably be isolated and purified.

The invention also provides an isolated and purified polypeptide having Swiss-prot data base Accession Number P23094 that comprises an amino acid other than phenylalanine at position 352, or a catalytically active variant, or a catalytically active fragment thereof. Preferably, the amino acid at position 352 is a naturally occurring amino acid. More preferably, the polypeptide is SEQ ID NO:2, 32, 33, 34, 35, or 36.

The invention also provides an isolated and purified NDO related polypeptide wherein the amino acid at the position corresponding to position 352

in NDO has been replaced with another amino acid, or a catalytically active fragment or catalytically active variant thereof. Preferably, the amino acid at the position corresponding to position 352 in NDO is a naturally occurring amino acid. More preferably, the amino acid at the position corresponding to position 352 in NDO is valine.

Site-directed mutagenesis was also used to construct DNA molecules that encode NDO mutants having an amino acid substitution at position 201, 202, 260, 316, 351, 352, 358, 362, or 366. Accordingly, the invention provides an isolated and purified NDO wherein the amino acid at position 201, 202, 260, 316, 351, 352, 358, 362, or 366 has been replaced with another amino acid, or a catalytically active variant, or a catalytically active fragment thereof.

Changing the amino acid at position 352 in NDO from phenylalanine to valine provided an enzyme (SEQ ID NO:2, encoded by SEQ ID NO:1) that gives a change in the preferred absolute configuration of the 1,2-dihydroxy-1,2-dihydronaphthalene formed from naphthalene. This enzyme also gave a change in the regioselectivity of the products obtained from oxidation of biphenyl and phenanthrene

The invention also provides an isolated and purified DNA segment encoding a polypeptide of the invention, or a variant or fragment thereof.

The invention also provides a primer or probe having about 80% nucleic acid sequence identity with a DNA segment encoding a polypeptide of the invention, or a variant or fragment thereof.

The invention also provides an expression cassette comprising a promotor operably linked to a DNA segment encoding a polypeptide of the invention or a variant or fragment thereof.

The invention also provides a host cell, the genome of which is augmented by a DNA segment encoding a polypeptide of the invention, or a catalytically active variant or fragment thereof.

The invention also provides a method to produce a catalytically active polypeptide comprising culturing a host cell transformed with a DNA

segment encoding a polypeptide of the invention, or a catalytically active variant or fragment thereof, so that the host cell expresses the polypeptide, variant or fragment.

The invention also provides a method for preparing *cis*-naphthalene dihydrodiol (e.g. (-)-(1*S*,2*R*)-*cis*-naphthalene dihydrodiol) comprising contacting naphthalene with a polypeptide of the invention, or a catalytically active variant or fragment thereof.

The invention also provides a method for preparing *cis*-naphthalene dihydrodiol (e.g. (-)-(1*S*,2*R*)-*cis*-naphthalene dihydrodiol) comprising contacting a host cell of the invention with naphthalene.

The invention also provides a method for preparing *cis*-biphenyl-3,4-dihydrodiol (e.g. (-) or (+) *cis*-biphenyl-3,4-dihydrodiol) comprising contacting biphenyl with a polypeptide of the invention, or a catalytically active variant or fragment thereof. The method may optionally further comprise dehydrating the *cis*-biphenyl-3,4-dihydrodiol to give 4-hydroxybiphenyl.

The invention also provides a method for preparing *cis*-biphenyl-3,4-dihydrodiol (e.g. (-)-*cis*-biphenyl-3,4-dihydrodiol) comprising contacting a host cell of the invention with biphenyl. The method may optionally further comprise dehydrating the (-)-*cis*-biphenyl-3,4-dihydrodiol to give 4-hydroxybiphenyl.

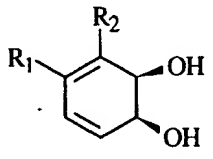
The invention also provides a method for preparing *cis*-phenanthrene-1,2-dihydrodiol (e.g. *cis*-(1S,2R)-phenanthrene-1,2-dihydrodiol) comprising contacting phenanthrene with a polypeptide of the invention, or a catalytically active variant or fragment thereof.

The invention also provides a method for preparing *cis*-phenanthrene-1,2-dihydrodiol (e.g. *cis*-(1S,2R)-phenanthrene-1,2-dihydrodiol) comprising contacting a host cell of the invention with phenanthrene.

The invention also provides a method to oxidize an aromatic compound to a corresponding dihydrodihydroxy compound comprising contacting the aromatic compound with a polypeptide of the invention, or a catalytically active variant or fragment thereof.

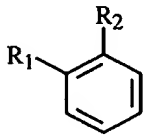
The invention also provides a method to oxidize an aromatic compound to a corresponding dihydrodihydroxy compound comprising contacting the aromatic compound with a host cell of the invention.

The invention also provides a method to prepare an optically active *cis*-cyclohexadiene of formula (I):



(I)

wherein one of R_1 and R_2 is hydrogen and the other is phenyl, 2-phenylvinyl, 2-phenylethynyl, or vinyl, wherein any phenyl ring may optionally be substituted with 1, 2, or 3 substituents independently selected from the group consisting of hydroxy, halo, carboxy, cyano, nitro, trifluoromethyl, amino, (C_1-C_6)alkyl, (C_1-C_6)alkoxycarbonyl, and (C_1-C_6)alkoxy, comprising contacting a corresponding compound of formula (II):



(II)

with a polypeptide of the invention, or a catalytically active variant or fragment thereof, or with a host cell of the invention. Preferably, R_1 is phenyl or 4-hydroxyphenyl. The method may optionally comprise dehydrating the resulting compound of formula I.

The invention also provides a method to prepare *cis*-1,2-dihydroxyindan comprising contacting indene with a polypeptide of the invention, or a catalytically active variant or fragment thereof, or with a host cell of the invention.

The invention also provides a method to prepare 1,2-dihydroxy-1,2,3,4-tetrahydronaphthalene comprising contacting 1,2-dihydronaphthalene with a polypeptide of the invention, or a catalytically active variant or fragment thereof, or with a host cell of the invention.

The invention also provides a method to prepare 1,2-dihydroxy-1,2-dihydrophenanthrene or 3,4-dihydroxy-3,4-dihydrophenanthrene comprising contacting phenanthrene with a polypeptide of the invention, or a catalytically active variant or fragment thereof, or with a host cell of the invention.

The invention also provides novel compounds and intermediates disclosed herein, as well as crystallized forms of the polypeptides disclosed herein. Preferably, the invention provides a crystallized form of SEQ ID NO:26.

Polypeptides of the invention provide a biosynthetic route to the either enantiomer of *cis*-naphthalene dihydrodiol, to either enantiomer of *cis*-biphenyl-3,4-dihydrodiol, and to either enantiomer of *cis*-phenanthrene-1,2-dihydrodiol. These products can be used in the synthesis of new polymers and pharmaceutical products. For example, arene *cis*-diols are useful starting materials for stereospecific organic synthesis (S.M. Brown, et al., *Organic Synthesis: Theory and Applications* (Hudlicky, T., ed.), pp. 113-176, JAI Press, Greenwich, CT (1993); and T. Hudlicky and J.W. Reed, *Adv. Asymm. Synth.* **1995**, *1*, 271-312). They can undergo a variety of reactions including asymmetric Diels-Alder reactions, epoxidation, photochemical oxygenation, metallation, diol cleavage, diene cleavage, carbene additions and ozonolysis. They have been used to prepare a variety of synthetic products that are not readily obtainable by conventional chemical synthesis. Examples include conduritols, inositol phosphates, pinitol enantiomers, prostanoid and terpene synthons, and complex natural products such as (−)-zeylena and (+)-lycoricidine.

Add B'

Detailed Description

The following definitions are used, unless otherwise described: halo is fluoro, chloro, bromo, or iodo. Alkyl, alkoxy, denote both straight and branched groups; but reference to an individual radical such as "propyl" embraces only the straight chain radical, a branched chain isomer such as "isopropyl" being specifically referred to.

The term "amino acid," comprises the residues of the natural occurring amino acids (e.g. Ala, Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Hyl, Hyp, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, and Val) in D or L form, as

well as unnatural amino acids (e.g. phosphoserine, phosphothreonine, phosphotyrosine, hydroxyproline, gamma-carboxyglutamate; hippuric acid, octahydroindole-2-carboxylic acid, statine, 1,2,3,4,-tetrahydroisoquinoline-3-carboxylic acid, penicillamine, ornithine, citruline, α-methyl-alanine, para-benzoylphenylalanine, phenylglycine, propargylglycine, sarcosine, and tert-butylglycine).

As used herein, the terms "isolated and/or purified" refer to *in vitro* preparation, isolation and/or purification of a nucleic acid molecule, sequence or segment of the invention, so that it is not associated with *in vivo* substances. Thus, with respect to an "isolated nucleic acid molecule sequence or segment", which includes a polynucleotide of DNA or RNA or of synthetic origin or some combination thereof, the "isolated nucleic acid molecule sequence or segment" (1) is not associated with all or a portion of a polynucleotide in which the "isolated nucleic acid molecule" is found in nature, (2) can be operably linked to a polynucleotide which it is not linked to in nature, or (3) does not occur in nature as part of a larger sequence. An isolated nucleic acid molecule means a polymeric form of nucleotides of at least 10 bases in length, either ribonucleotides or deoxynucleotides or a modified form of either type of nucleotide. The term includes single and double stranded forms of DNA. The term "oligonucleotide" referred to herein includes naturally occurring, and modified nucleotides linked together by naturally occurring, and non-naturally occurring oligonucleotide linkages. Oligonucleotides are a polynucleotide subset with 200 bases or fewer in length. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19, or 20 to 40 bases in length. Oligonucleotides are usually single stranded, e.g., for probes or primers; although oligonucleotides may be double stranded, e.g., for use in the construction of a variant. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. The term "naturally occurring nucleotides" referred to herein includes deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "oligonucleotide linkages" referred to herein includes oligonucleotides linkages such as phosphorothioate,

phosphorodithioate, phophoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. An oligonucleotide can include a label for detection, if desired.

The term "isolated polypeptide" means a polypeptide encoded by DNA or RNA, including polypeptides that are synthetic in origin, or some combination thereof, which isolated polypeptide (1) is not associated with proteins found in nature, (2) is free of other proteins from the same source, (3) is expressed by a cell from a different species, or (4) does not occur in nature.

The term "selectively hybridize" means to detectably and specifically bind. Polynucleotides, oligonucleotides and fragments of the invention selectively hybridize to nucleic acid strands under hybridization and wash conditions that minimize appreciable amounts of detectable binding to nonspecific nucleic acids. High stringency conditions can be used to achieve selective hybridization conditions as known in the art and discussed herein. Generally, the nucleic acid sequence identity between the polynucleotides, oligonucleotides, variants, and fragments of the invention and a nucleic acid sequence of interest is at least about 65%, and more typically with preferably increasing identities of at least about 70%, about 90%, about 95%, about 98%, and 100%. See Sambrook et al. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor (1989).

The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (i.e., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence, or that a polypeptide sequence is identical to a reference polypeptide sequence. In contradistinction, the term "complementary to" is used herein to mean that the complementary sequence is homologous to all or a portion of a reference polynucleotide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

The term "NDO related polypeptide (or complex)" means a polypeptide (or complex) that belongs to the same family of bacterial enzymes as NDO or TDO. Preferably, an NDO related polypeptide has substantial identity

with NDO. More preferably, an NDO related polypeptide is encoded by the DNA sequence having Genbank Accession number M60405, M23914, AF010471, AF004284, M83949, AF004283, AB004059, D84146, AF036940, U49504, or U62430.

The term "catalytically active," when applied to a polypeptide or a polypeptide variant or fragment of the invention, means that the variant or fragment catalyzes one or more of the reactions catalyzed by NDO or an NDO related polypeptide (see for example S.M. Resnick, et al., *Journal of Industrial Microbiology*, **1996**, *17*, 438-457). Preferably, a catalytically active polypeptide, or a catalytically active variant, or fragment catalyzes one or more of the reactions catalyzed by NDO. More preferably, a catalytically active polypeptide, or a catalytically active variant, or fragment catalyzes the oxidation of an aromatic substrate to give the corresponding dihydrodihydroxy compound (e.g. the oxidation of biphenyl or phenanthrene to 3,4-dihydroxy-3,4-dihydrobiphenyl or 1,2-dihydroxy-1,2-dihydrophenanthrene, respectively).

The term "biologically active" with respect to a fragment or variant of a polypeptide means that the fragment or variant has at least about 10%, preferably at least about 50%, and most preferably at least about 90%, the activity of the reference polypeptide. The activity of a polypeptide of the invention can be measured by methods well known to the art including, but not limited to, the ability of the peptide to elicit a sequence-specific immunologic response when the peptide is administered to an organism, e.g., chicken, goat, sheep or mice. The invention also provides polypeptides, as well as fragments and variants thereof, that are useful to elicit a sequence-specific immunologic response when the peptide is administered to an organism (e.g. a mammal).

When applied to a polypeptide the term "fragment" means a portion of the reference polypeptide that comprises the amino acid that corresponds to amino acid 352 in SEQ ID NO:26. Preferably the portion of the reference polypeptide also comprises at least about 10, 50, or 500 additional amino acids of the reference peptide. More preferably, the portion of the reference polypeptide comprises at least about 1000 or 1500 additional amino acids or the reference peptide.

When applied to a nucleotide sequence the term "fragment" means a portion of the reference nucleotide sequence that 1) encodes the amino acid corresponding to amino acid 352 in SEQ ID NO:26, and 2) encodes a catalytically active polypeptide. Preferably the portion of the reference nucleotide sequence also comprises at least about 30, 60, 150, or 300, nucleoside bases of the reference nucleotide sequence. More preferably, the portion of the reference nucleotide sequence also comprises at least about 600, 900, or 1200 nucleoside bases of the reference nucleotide sequence.

The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity".

A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length DNA or gene sequence given in a sequence listing, or may comprise a complete DNA or gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by the local homology algorithm of Smith and

Waterman (1981) *Adv. Appl. Math.* 2: 482, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol. 48:* 443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. (U.S.A.) 85*: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. "Percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denote a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least about 85 percent sequence identity, preferably at least about 90 to about 95 percent sequence identity, more usually at least about 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 20-50 or 50-200 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison.

As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least about 80 percent sequence identity, preferably at least about 80 percent sequence identity, more

preferably at least about 90 percent sequence identity, and most preferably at least about 99 percent sequence identity. Alternatively two polypeptide sequences have substantial identity if they have an alignment score of at more than 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. See Dayhoff, M. O., in Atlas of Protein Sequence and Structure, 1972, volume 5, National Biomedical Research Foundation, pp. 101-110, and Supplement 2 to this volume, pp. 1-10. Preferably, the two sequences have substantial identity if their amino acids are greater than or equal to 50% identical when optimally aligned using the ALIGN program. When a percent sequence identity is given, it means that the stated percentage of the amino acids are identical when the two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less are preferred with 2 or less being more preferred.

As used herein, "substantially pure" means an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual species in the composition), and preferably a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 percent of all macromolecular species present in the composition, more preferably more than about 85%, about 90%, about 95%, and about 99%. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species.

When applied to a polypeptide the term "variant," means a polypeptide that 1) has substantial identity with but is not identical to the reference polypeptide; and 2) is identical to the reference polypeptide at the position corresponding to amino acid 352 in SEQ ID NO:26.

When applied to a nucleotide sequence, the term "variant" means a nucleotide sequence that 1) has substantial identity with but is not identical to

the reference sequence; and 2) encodes the same amino acid as the reference sequence at the position encoded by the reference sequence that corresponds to amino acid 352 in SEQ ID NO:26.

The variant DNA molecules of the invention may include DNA molecules with "silent" substitutions. For example, a preferred embodiment of the invention is an isolated and purified DNA molecule comprising a preselected DNA segment encoding SEQ ID NO:2, wherein the DNA segment comprises SEQ ID NO:1, or variants of SEQ ID NO:1, having nucleotide substitutions which are "silent" (see Table 1). That is, when silent nucleotide substitutions are present in a codon, the same amino acid is encoded by the codon with the nucleotide substitution as is encoded by the codon without the substitution. "Silent" nucleotide substitutions in SEQ ID NO:1 which can encode a peptide having SEQ ID NO:2 can be ascertained by reference to Table 1 and page D1 in Appendix D in Sambrook et al., Molecular Cloning: A Laboratory Manual (1989). Nucleotide substitutions can be introduced into DNA segments by methods well known to the art. See, for example, Sambrook et al., *supra*.

Table 1

Amino Acid	Codon
Phe	UUU, UUC
Ser	UCU, UCC, UCA, UCG, AGU, AGC
Tyr	UAU, UAC
Cys	UGU, UGC
Leu	UUA, UUG, CUU, CUC, CUA, CUG
Trp	UGG
Pro	CCU, CCC, CCA, CCG
His	CAU, CAC
Arg	CGU, CGC, CGA, CGG, AGA, AGG
Gln	CAA, CAG
Ile	AUU, AUC, AUA
Thr	ACU, ACC, ACA, ACG
Asn	AAU, AAC
Lys	AAA, AAG

Amino Acid	Codon
Met	AUG
Val	GUU, GUC, GUA, GUG
Ala	GCU, GCC, GCA, GCG
Asp	GAU, GAC
Gly	GGU, GGC, GGA, GGG
Glu	GAA, GAG

Specific values listed below for radicals, substituents, and ranges, are for illustration only; they do not exclude other defined values or other values within defined ranges for the radicals and substituents.

Specifically, $(C_1\text{-}C_6)$alkyl can be methyl, ethyl, propyl, isopropyl, butyl, iso-butyl, sec-butyl, pentyl, 3-pentyl, or hexyl; and $(C_1\text{-}C_6)$alkoxy can be methoxy, ethoxy, propoxy, isopropoxy, butoxy, iso-butoxy, sec-butoxy, pentoxy, 3-pentoxy, or hexyloxy.

Pseudomonas sp NCIB 9816-4 contain an inducible multicomponent enzyme system designated naphthalene dioxygenase (NDO) which catalyzes the formation of *cis*-naphthalene dihydrodiol. The system consists of an iron-sulfur flavoprotein ($reductase_{NAP}$), a Rieske [2Fe-2S] protein ($ferredoxin_{NAP}$), and an iron-sulfur protein (ISP_{NAP}), which serves as the terminal oxygenase component. ISP_{NAP} has an $\alpha_3\beta_3$ subunit component and each α subunit contains a Rieske [2Fe-2S] cluster and mononuclear iron. The Rieske cluster is believed to be an electron storage center that transfers electrons to mononuclear iron which is responsible for dioxygen activation and ultimately the catalytic reaction.

The genes encoding the NDO complex in *Pseudomonas* sp NCIB 9816-4 have been cloned and expressed in *Escherichia coli* (W-C Suen and D.T. Gibson, *Gene*, **1994**, *143*, 67-71; and W-C Suen Ph.D. Thesis The University of Iowa, Iowa City, Iowa, **1993**). The nucleotide sequences of the genes encoding $reductase_{NAP}$ (*nahAa*), $ferredoxin_{NAP}$ (*nahAb*), and ISP_{NAP} (*nahAcAd*) have been determined and show 93.3%, 93.3%, 96.9%, and 94.8% identity, respectively, at the predicted amino acid level with the isofunctional genes carried by the well-

studied NAH7 plasmid in *P. putida* G7 (S. Kurkele, *Gene*, **1988**, *73*, 355-362; M.J. Simon, *Gene*, **1993**, 127, 31-37).

A number of dioxygenases with a structure similar to NDO have been identified. For example, the dioxygenases shown in Table 2 have α subunits with greater than 80% amino acid identity to naphthalene dioxygenase from *Pseudomonas* sp. strain NCIB 9816-4 (calculated using GAP Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wisconsin). Because of their similarity to NDO in structure and function, it is reasonable to believe that modifying these polypeptides by replacing the amino acid corresponding to F352 in NDO with valine, will provide novel polypeptides that have similar advantageous properties compared to the native polypeptides. Accordingly, the invention also provides the following DNA molecules (SEQ ID No.'s 3-13) that have been modified to encode valine at the position corresponding to the F352 amino acid in NDO, as shown in Table 2.

Table 2

Genbank Accession Number	Amino Acid Corresponding to F352 in NDO	SEQ ID NO. For Modified DNA Sequence	SEQ ID NO. For Corresponding Polypeptide
M60405	F352	SEQ ID NO:3	SEQ ID NO:14
M23914	F352	SEQ ID NO:4	SEQ ID NO:15
AF010471	F352	SEQ ID NO:5	SEQ ID NO:16
AF004284	F352	SEQ ID NO:6	SEQ ID NO:17
M83949	F352	SEQ ID NO:7	SEQ ID NO:18
AF004283	F352	SEQ ID NO:8	SEQ ID NO:19
AB004059	F352	SEQ ID NO:9	SEQ ID NO:20
D84146	F352	SEQ ID NO:10	SEQ ID NO:21
AF036940	F350	SEQ ID NO:11	SEQ ID NO:22
U49504	I350	SEQ ID NO:12	SEQ ID NO:23
U62430	T355	SEQ ID NO:13	SEQ ID NO:24

The invention also provides the polypeptides (SEQ ID No.'s 14-24) that are encoded by the DNA molecules of SEQ ID No.'s 3-13. The invention also provides a host cell, the genome of which is augmented by a DNA molecule having SEQ ID NO:3-13.

In addition to the enantiospecific *cis*-dihydroxylation of naphthalene, NDO catalyzes dioxygenation of a variety of multicyclic and heterocyclic aromatic compounds to produce, in many cases, chiral *cis*-dihydrodiols. The mutant and recombinant strains expressing polypeptides of the invention allow the synthesis of *cis*-diols in high yields and/or high enantiomeric purity. For example, substrates which are oxidized to *cis*-dihydrodiols by NDO include indene, 1,2-dihydronaphthalene, benzocyclohept-1-ene, anthracene, phenanthrene, dibenzo[1,4]dioxan, acenaphthylene, 1- and 2-substituted naphthalenes, biphenyl, fluorene, dibenzofuran, dibenzothiophene, 9,10-dihydroanthracene, and 9,10-dihydrophenanthrene.

The DNA segment having Genbank Accession Number U49496 is SEQ ID NO:25. The polypeptide having Swiss-prot data base Accession Number P23094 is SEQ ID NO:26.

The invention will now be illustrated by the following non-limiting Examples.

Examples

Example 1. **Construction of site-directed mutations.**

Escherichia coli strains DH5α (Life Technologies, Gaithersburg, MD) and JM109(DE3)) (Promega Corp., Madison, WI) were used for subcloning and gene expression experiments, respectively. Competent *E. coli* strains ES1301 and JM109 were purchased from Promega Corp. and used in the site-directed mutagenesis procedure described below.

E. coli strains were grown at 30 or 37 °C in Luria-Bertani (LB) medium (R.W. Davis, et al. Advanced Bacterial Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1980), or Terrific Broth (TB) medium (Lee, S.-Y., and S. Rasheed, *BioTechniques,* **1990**, 9, 676-679). To maintain plasmids, ampicillin or tetracycline was added to final concentrations of 150 and

12 µg/ml, respectively. JM109(DE3) strains carrying plasmids of interest were maintained on minimal medium plates (MSB) (R.Y. Stanier, et al., *J. Gen. Microbiol.*, **1966**, *43*, 159-271) containing 10 mM glucose, 0.1 mM thiamine, and ampicillin. For plates, MSB was solidified with 1.8% Agar Noble (Difco Laboratories, Detroit, Mich.) and LB was solidified with 1.5% Bactoagar (Difco Laboratories).

Mutagenesis of *nahAc* (encoding the naphthalene dioxygenase α subunit) was carried out with the Altered Sites II *in vitro* Mutagenesis System according to the manufacturer's instructions (Promega Corp.). A 1.5-kb *Kpn*I-*Xba*I fragment carrying the 3' half of the *nahAc* gene and the complete *nahAd* gene from pDTG141 (Suen, W.-C, Gene expression of naphthalene dioxygenase from *Pseudomonas* sp. NCIB 9816-4 in *Escherichia coli*. Ph. D. thesis. The University of Iowa, Iowa City, Iowa, 1991) was cloned into *Kpn*I-*Xba*I-digested pALTER-1 (Promega Corp.). The resulting plasmid, designated pMASTER-1, was used as the template for mutagenesis.

The mutagenic oligonucleotide (5'-TTCAGCGAACGGTCGG-GCCTGC-3') (SEQ ID NO:37) was designed such that the restriction pattern of the plasmid was altered (eliminating a *Psp*1406I site shown by underlined bases; T-G base change shown in bold) to facilitate screening for clones carrying the desired mutation. The same T-G base change alters the TTC codon specifying phenylalanine to GTC, which specifies valine.

The phosphorylated oligonucleotide used for mutagenesis was synthesized by Genosys Biotechnologies Inc., Midland, Tex. pMASTER-1 (2 µg) was denatured at room temperature for 5 min in the presence of 200 mM NaOH, 0.2 mM ethylenediamine tetraacetic acid (EDTA) in a 20 µl volume. The denatured plasmid DNA was precipitated by addition of 2 M ammonium acetate (2 µl; pH 4.6) and 100% ethanol (75 µl) and incubated at -70 °C for 30 minutes. After centrifugation for 15 minutes at 14,000 rpm in an Eppendorf centrifuge, the DNA pellet was dried under vacuum and dissolved in 10 mM Tris, 1 mM EDTA (100 ml; pH 8.0).

The primer annealing reaction was carried out with the following components in a final volume of 20 µl: denatured pMASTER-1 (10 µl);

tetracycline repair oligonucleotide (1 µl; 5'-GCCGGGCCTCTTGCGGG-ATATCGTCCA-3') (SEQ ID NO:38); ampicillin knockout oligonucleotide (1 µl; 5'-GTTGCCATTGCTGCAG-GCATCGTGGTG-3') (SEQ ID NO:39); phosphorylated mutagenic oligonucleotide 1.25 pmol); 10X annealing buffer (2 µl; 200 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 500 mM NaCl).

The mixture was heated at 75 °C for 5 minutes and cooled slowly (1°C per min) to 45 °C, then more rapidly to room temperature. The synthesis reaction mix contained the products of the annealing reaction and the following: sterile distilled water (5 µl); 10X synthesis buffer (3 µl, 100 mM Tris-HCl (pH 7.5), 5 mM deoxyribonucleotides, 10 mM adenosine triphosphate, 20 mM dithiothreitol); T4 DNA polymerase (1 µl); T4 DNA ligase (1 µl).

The mixture was incubated at 37°C for 90 minutes. A portion of this mixture (1.5 µl) was used to transform 100 µl of ES1301*mutS* competent cells. The transformation mix was incubated on ice for 10 minutes, heated at 42°C for 45 seconds and then incubated on ice for 2 minutes. LB (900 µl) was added and the culture was incubated at 37°C for 30 minutes without shaking. After incubation, 500 µl of this culture was diluted with 4.5 ml of LB containing 1.25 µg/µl ampicillin and incubated over night at 37°C with shaking. Plasmid DNA was purified from the overnight culture as described by Lee, S.-Y., and S. Rasheed, *BioTechniques,* **1990**, 9, 676-679, and was used to transform JM109 using standard procedures similar to those described by F.M. Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, Inc., New York, · 1993.

The transformation mixture was plated on LB plates containing ampicillin. Colonies were screened for tetracycline sensitivity on LB plates containing 12 µg/ml tetracycline. Plasmid DNA purified from ampicillin resistant, tetracycline sensitive strains was digested with *AclI* (New England Biolabs, Beverly, MA) and the restriction pattern was compared to that of unmutagenized pMASTER-1 using standard molecular biology methods (for example see F.M. Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, Inc., New York, 1993). The nucleotide sequences of both strands

of the entire insertion in pMASTER-1 were determined for one mutant plasmid that contained the restriction site change.

Fluorescent automated DNA sequencing was carried out by the University of Iowa DNA Facility using an Applied Biosystems 373A automated DNA sequencer. After verification of each mutation by restriction digestion and sequence analysis, the 1.5-kb *Kpn*I-*Xba*I fragments carrying each mutation were individually cloned into *Kpn*I-*Xba*I-digested pDTG141 (which contains the *nahAaAbAcAd* genes for the naphthalene dioxygenase reductase, ferredoxin and α and β subunits of the oxygenase, respectively. The resulting plasmids were introduced into JM109(DE3) for expression studies. After this subcloning step, the presence of the mutation was verified by restriction analysis and one sequencing run with a primer that generated sequence in the region of the mutagenized base and continued past the *Kpn*I junction.

Example 2. **Biotransformations.**

To produce induced cells for biotransformation studies, JM109(DE3) carrying the plasmid (pDTG141-F352V) [F=phenylalanine; V=valine] was grown in flasks at 30°C in minimal medium (MSB) (R.Y. Stanier, et al., *J. Gen. Microbiol.*, **1966**, *43*, 159-271) containing 10 mM glucose, 0.1 mM thiamine, and 150 μg/ml ampicillin with shaking (200 rpm). JM109(DE3)(pDTG141) was grown under identical conditions to provide the wild-type control. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 100 μM when culture turbidity reached 0.6-0.8 at 660 nm. After a 2.5 hour induction period, biotransformation reactions were initiated. Cultures were spiked with 20 mM glucose, 100 mM phosphate buffer (pH 7.2), and 0.025% (w/v) substrate (e.g. naphthalene, biphenyl, phenanthrene). Cultures were incubated at 30°C with shaking (250 rpm) for up to 18 hours. Samples were taken periodically and cells were removed by centrifugation.

Culture supernatants were extracted with sodium hydroxide-washed ethyl acetate and analyzed by thin layer chromatography (S.M. Resnick, et al., *FEMS Microbiol. Lett.*, **1993**, *113*, 297-302). All extracts were analyzed by gas chromatography-mass spectrometry (GC-MS) as previously described

(S.M. Resnick, S. *Appl. Environ. Microbiol.*, **1994**, *60*, 3323-3328). In some cases, dihydrodiol products were derivatized with approximately equimolar amounts of phenyl boronic acid at room temperature prior to injection onto the GC column.

Generally, biotransformation products were most easily detected by thin-layer chromatography (TLC), high-performance liquid chromatography (HPLC), and/or gas chromatography-mass spectrometry (GC/MS). Preparative TLC, HPLC, column or radial-dispersion chromatography are commonly used for the isolation of oxidation products prior to nuclear magnetic resonance (NMR) structural determination and/or stereochemical analysis.

Results from the biotransformation studies are shown in Table 3.

Table 3

NDO Polypeptide	Enantiomeric Composition of cis-naphthalene 1,2-dihydrodiol product	Dihydrodiols From Biphenyl		Dihydrodiols From Phenanthrene	
		%2,3	%3,4	%3,4	%1,2
Wild Type	>99	87.2	12.8	91.3	8.7
352-valine	92.4	4.4	95.6	17.0	83.0

In the biotransformations described above, incubation of biphenyl with host cells carrying the plasmid (pDTG141-F352V) yielded predominantly 3,4-dihydroxy-3,4-dihydrobiphenyl. This is in contrast to results obtained with the wild type cells, which yielded predominantly the 2,3-dihydroxy-2,3-dihydrobiphenyl. Additionally, incubation of phenanthrene with host cells carrying the plasmid (pDTG141-F352V) yielded predominantly 1,2-dihydroxy-1,2-dihydrophenanthrene, rather than the corresponding 3,4 isomer preferentially formed from the wild type cells.

Thus, the polypeptides of the invention and the host cells of the invention are useful for preparing chiral diols for use in the polymer, resin, pharmaceutical or rubber industry. In particular, the polypeptides of the invention and the host cells of the invention are useful for preparing (-)-*cis*-3,4-

dihydroxy-3,4-dihydrobiphenyl, and a single enantiomer (1S,2R) of *cis*-1,2-dihydroxy-1,2-dihydrophenanthrene.

The compound (-)-*cis*-3,4-dihydroxy-3,4-dihydrobiphenyl can be dehydrated to provide 4-hydroxybiphenyl, which is useful for the manufacture of rubber and resins (see The Merck Index (Martha Windholz, ed.), 10 ed., 7187 (*p*-phenylphenol) Merck & Co. Inc. New Jersey, USA). Thus, the invention also provides an environmentally benign route to this useful compound. Additionally, the polypeptides of the invention and the host cells of the invention may be useful for carrying out bioremediation.

Example 3. **(+)-*cis*-(1*R*,2*S*)-Dihydroxy-1,2-dihydronaphthalene**

cis-Dihydroxy-1,2-dihydronaphthalene formed by naphthalene dioxygenase was purified for chiral HPLC analysis by preparative-layer chromatography (A.M. Jeffrey, et al. *Biochemistry*, **1975**, *14*, 575-583; S.M. Resnick, S. *Appl. Environ. Microbiol.*, **1994**, *60*, 3323-3328). Chiral stationary-phase liquid chromatography (CSP-HPLC) was used to resolve the two enantiomers of *cis*-naphthalene dihydrodiol with a Chirocel OJ column (Chiral Technologies, Exton, PA) as described by S.M. Resnick, S. *Appl. Environ. Microbiol.*, **1994**, *60*, 3323-3328. Under these conditions, the (+)-(1*R*,2*S*) and (-)-(1*S*,2*R*) enantiomers of *cis*-naphthalene dihydrodiol eluted with retention times of 30 and 33 minutes, respectively.

Example 4. **(-)-*cis*-Biphenyl-3,4-dihydrodiol.**

A 6 L culture of JM109(DE3)(pDTG141-F352V) was grown in a 10 L Biostat B fermentor (B.Braun Biotech International, Melsungen, Germany) in MSB at 27 °C. Automated addition of NH_4OH was used to maintain the pH at 7.3, and a slow glucose feed was used to maintain dissolved O_2 concentration at approximately 25% saturation. The culture was induced for 3 hours with 150 μM IPTG when the optical density of the culture (660 nm) reached 0.8. After 17 hours incubation with 0.025% (w/v) biphenyl, cells were harvested by centrifugation and the culture supernatant was extracted with ethyl acetate and concentrated as described previously (S.M. Resnick, et al., *FEMS Microbiol.*

Lett., **1993**, *113*, 297-302). Two purification methods were employed: 1) multiple elution preparative thin layer chromatography with a 95:5 mixture of chloroform and acetone as eluting solvent as previously described (D.S. Torok, et al., *J. Bacteriol.*, **1995**, *177*, 5799-5805); and 2) radial dispersion chromatography eluting with a step gradient of chloroform-methanol (S.M. Resnick, S. *Appl. Environ. Microbiol.*, **1994**, *60*, 3323-3328).

Over 150 mg of biphenyl 3,4-dihydrodiol was obtained from 500 mg crude product. The purity of the compound was determined by thin layer chromatography. The positions of the hydroxyl groups on the aromatic ring were determined by ^{1}H NMR analysis. The enantiomeric purity was determined to be >97% by subjecting the corresponding phenyl-boronate derivative to gas-chromatography-mass spectrometry analysis (see S.M. Resnick et al., *J. Org. Chem.*, **1995**, *60*, 3546-3549). Physical properties of the compound were as follows: λmax (in methanol) 276 and 228 nm, ϵ_{276}=4,336 ± 574 and ϵ_{228}=18,580 ± 1621; calculated mass of the phenyl boronate derivative ($C_{18}H_{15}O_2B$) was 274.1165, found mass, 274.1160; $[\alpha]_D$ -37.5 ±3.8 (c=0.5 g/100 mL, methanol) as determined using a Jasco P1020 polarimeter with a Na 589 nm lamp.

Using a procedure similar to that described in Example 3, *cis*-biphenyl 2,3-dihydrodiol and *cis*-biphenyl 3,4-dihydrodiol were separated following multiple elution preparative thin layer chromatography. Under the same chiral HPLC conditions used to separate enantiomers of *cis*-naphthalene dihydrodiols, the (+)- and (-)- enantiomers of *cis*-biphenyl 3,4-dihydrodiol eluted with retention times of 31 and 28 minutes, respectively. The (+)-(2*R*,3*S*) and (-)-(2*S*,3*R*) enantiomers of *cis*-biphenyl 2,3-dihydrodiol eluted with retention times of 28 and 33 min, respectively.

Example 5. **4-Hydroxybiphenyl**

Acid-catalyzed dehydration of biphenyl 3,4-dihydrodiol gave predominantly 4-hydroxybiphenyl. The addition of 50 mM HCl (final concentration) to a 0.3 mM solution of biphenyl 3,4-dihydrodiol in methanol at room temperature gave complete dehydration in minutes.

Example 6. **Mutations in NDO**

Bacterial strains and plasmids. Bacterial strains and plasmids used in this study are listed in **Table 4**. *Escherichia coli* strains DH5α and JM109(DE3) were used for subcloning and gene expression experiments, respectively. Competent *E. coli* strains ES1301 and JM109 were purchased from Promega Corp., Madison, Wis. and used in the site-directed mutagenesis procedure described below.

Media and growth conditions. *E. coli* strains were grown at 37°C in Luria-Bertani (LB) medium (12), or Terrific Broth (TB) medium. Antibiotics were added to the following final concentrations as appropriate: ampicillin, 150 μg/ml; tetracycline, 12.5 μg/ml. To produce induced cells for biotransformation studies, JM109(DE3) strains carrying plasmids of interest were grown at room temperature 30°C in minimal salts medium (MSB) containing 10 mM glucose, 0.1 mM thiamine, and ampicillin. Isopropyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 100 μM when culture turbidity reached 0.6-0.8 at 660 nm. After a 2 hour induction, biotransformations were initiated as described below. For plates, MSB was solidified with 1.8% Noble Agar (Difco Laboratories, Detroit, Mich.) and LB was solidified with 1.5% Bactoagar (Difco Laboratories).

Molecular techniques. Plasmid DNA was isolated as described previously (S-Y Lee, S. Rasheed, *Biotechniques*, **1990**, *9*, 676-679) or by using the Qiagen Midi Kit (Qiagen, Inc., Chatsworth, Calif.). For nucleotide sequencing, DNA was further purified using a Centricon-100 filter unit (Amicon, Inc., Beverly, Mass.). Restriction digests were performed as suggested by the enzyme suppliers (New England Biolabs, Inc., Beverly, Mass.; Promega Corp., Madison, Wis.). DNA fragments were purified from gel slices using the GeneClean Spin Kit according to the manufacturer's instructions (BIO101, Vista, Calif.). Ligation reactions, transformation of *E. coli* strains and agarose gel electrophoresis were performed by standard procedures.

TABLE 4. Strains and plasmids used in this study

Strain or plasmid	Relevant Characteristics[a]	Source or Reference
***E. coli* strains**		
DH5α	Δ(*lacZYA-argF*)U169, *hdsR*17 *relA*1, *supE*44, *endA*1, *recA*1, *thi gyrA*96, ϕ80d*lacZ*ΔM15	Life Technologies, Gaithersburg, MD
JM109	*endA*1, *recA*1, *gyrA*96 *thi*, *hdsR*17 *relA*1, *supE*44, Δ(*lac-proAB*), *mcrA*, [F', *traD*36, *proAB*$^+$, *lacI*^{q}ZΔM15]	C. Yanisch-Perron et al., *Gene*., 1985 *33*:103-119.
JM109(DE3)	*endA*1, *recA*1, *gyrA*96 *thi*, *hdsR*17 *relA*1, *supE*44, Δ(*lac-proAB*), *mcrA*, [F', *traD*36, *proAB*$^+$, *lacI*^{q}ZΔM15], λ(DE3)	Promega Corp., Madison, Wis.
ES1301 *mutS*	Km^r, *lacZ*53, *mutS*201::Tn*5*, *thyA*36, *rha*-5, *metB*1, *deoC*, IN(*rrnD-rrnE*)	Promega Corp., Madison, Wis.
Plasmids		
pDTG141	Ap^r, *nahAaAbAcAd* (encoding the naphthalene dioxygenase components reductase$_{NAP}$, ferredoxin$_{NAP}$, and large and small subunits of the oxygenase, respectively) under the control of the T7 promoter of pT7-5	W.-C. Suen, 1991. Ph.D. Thesis. The University of Iowa, Iowa City, IA.
pMASTER-1	Tc^r, Ap^S, pALTER-1 carrying the *Kpn*I-*Xba*I fragment of pDTG141 (*nahAc'Ad*)	R. E. Parales et al., *J. Bacteriol.* 1999, *181*:1831-1837.

[a] Km^r, kanamycin resistance; Ap^r, ampicillin resistance; Tc^r, tetracycline resistance

Site-directed mutagenesis. Mutagenesis of *nahAc* was carried out with the Altered Sites II *in vitro* Mutagenesis System according to the manufacturer's instructions (Promega Corp., Madison, Wis.). Plasmid pMASTER-1 (R.E. Parales et al., *J. Bacteriol.*, **1999**, *181*, 1831-1837), which contains the 3' end of the *nahAc* gene and the complete *nahAd* gene (which encode the α and β subunits of NDO, respectively), was used as the template for mutagenesis. Each mutagenic oligonucleotide was designed with a silent mutation that altered the restriction pattern of the plasmid **(Table 5)** to facilitate screening for clones carrying the desired mutation. Phosphorylated oligonucleotides used for mutagenesis were synthesized by Genosys Biotechnologies Inc., Midland, Tex. The nucleotide sequences of both strands of the entire insertion in pMASTER-1 were determined for each mutant. Fluorescent automated DNA sequencing was carried out by the University of Iowa DNA Facility using an Applied Biosystems 373A automated DNA sequencer.

After verification of each mutation by restriction digestion and sequence analysis, the 1.5-kb *Kpn*I-*Xba*I fragments carrying each mutation were individually cloned into *Kpn*I-*Xba*I-digested pDTG141. After this subcloning step, the presence of each mutation was verified by restriction and sequence analyses. The resulting derivatives of pDTG141 were introduced into JM109(DE3) for expression studies and in this way each protein isoform was produced from an identical expression system.

Whole cell biotransformations. Induced *E. coli* cultures (50 ml) were supplemented with 20 mM glucose and 80 mM phosphate buffer (pH 7.2). Solid substrates (naphthalene, biphenyl, or phenanthrene) were added to a final concentration of 0.025% (w/v). Cultures were incubated at 30 °C with shaking (250 rpm) for 15-18 h. To obtain cells for large scale biotransformations to produce *cis*-biphenyl 3,4-dihydrodiol, JM109(DE3)(pDTG141-F352V) was grown at 27 °C in MSB containing glucose, thiamine, and ampicillin in a 10 L Biostat B fermentor (B. Braun Biotech International, Melsungen, Germany). Automated addition of NH_4OH was used to maintain the pH at 7.3, and a slow

glucose feed rate was used to maintain the dissolved O_2 concentration at approximately 25% saturation.

Cultures were induced for 3 hours with 150 μM IPTG when the optical density of the culture (660 nm) reached approximately 0.7. Induced cultures (5.5 L)were incubated at 27°C for 14-17 h with 0.025% (w/w) substrate (biphenyl or phenanthrene), high agitation (700 rpm), automated pH control (pH 7.5) and a slow glucose feed.

Indigo formation. JM109(DE3) strains carrying pDTG141 derivatives with the various mutations were grown overnight at 37 °C on nitrocellulose filters placed on MSB agar plates containing 10 mM glucose, 1 mM thiamine, and 150 μg/ml ampicillin. Dried Whatman #1 filter papers that had been soaked in a 10% solution of indole dissolved in acetone were placed in the Petri dish covers after colony formation. Production of indigo from indole vapor by NDO was observed as colonies turned blue. No induction was carried out for these studies.

Separation and identification of products. Culture supernatants from whole cell biotransformation experiments were extracted with sodium hydroxide-washed ethyl acetate and analyzed by thin layer chromatography (TLC). Phenyl boronic acid (PBA) derivatives (A.B. Herbert, European Patent EP 0379300A2) were prepared as previously described (S.M. Resnick, D.T. Gibson, *Appl. Environ. Microbiol.* **1996**, *62*, 4073-4080). PBA-derivatized extracts were analyzed by gas chromatography-mass spectrometry (GC-MS) as previously described (S.M. Resnick, D.T. Gibson, *Appl. Environ. Microbiol.* **1996**, *62*, 3355-3359). *cis*-Naphthalene dihydrodiol was purified by preparative-layer chromatography (PLC) with chloroform-acetone (8:2) (S.M. Resnick, et al., *Appl. Environ. Microbiol.* **1994**, *60*, 3323-3328).

TABLE 5. Amino acid substitutions in the α subunit of NDO generated by site-directed mutagenesis

Mutation	Mutagenic oligonucleotide[a]	Restriction site change	Indigo formation[b]
N201A	5'-GAGGCACCCGCGG**AAGCTT**TTGTGGGAGATGCA-3 (SEQ ID NO:40)	*Hin*dIII	+
N201Q	5'-GCACCCGCGG**AACAAT**TTGTGGGAGATGCA-3' (SEQ ID NO:41)	*Tsp*509I	+
N201S	5'-CCGCGGA**AAGCTT**TGTGGGAG-3' (SEQ ID NO:42)	*Hin*dIII	++
F202L	5'-CCGCGGAA**AAGCTT**GTGGGAGATG-3' (SEQ ID NO:43)	*Hin*dIII	-
F202V	5'-CGCGGAA**AACGTT**GTGGGAGATG-3' (SEQ ID NO:44)	*Acl*I	++
V260A	5'-ATATTCAGG**TGCGCA**TAGCGCAG-3' (SEQ ID NO:45)	*Fsp*I	++
V260L	5'-GGACGGATATTCA**GGGCTCC**ATAGCGCAGACTTG-3' (SEQ ID NO:46)	*Ban*II	++
V260N	5'-GACGGATATTCA**GGTAACC**ATAGCGCAGACTTG-3' (SEQ ID NO:47)	*Bst*EII	++
W316A	5'-GGTGTTTTCAAAG**TCGCGA**ACCCGATCGAC-3' (SEQ ID NO:48)	*Nru*I	+++
T351N	5'-CTGTTCAGCG**AAACTT**CGGGCCTGCT-3' (SEQ ID NO:49)	remove *Acl*I	++
T351R	5'-CTGTTCAGCG**AAGGTT**CGGGCCTGCT-3' (SEQ ID NO:50)	remove *Acl*I	+
T351S	5'-CTGTTCAGCG**AAGCTT**CGGGCCTGCT-3' (SEQ ID NO:51)	*Hin*dIII	+++
F352L	5'-TTCAGCG**AACGCT**CGGGCCTGC-3' (SEQ ID NO:52)	remove *Acl*I	++
F352V	5'-TTCAGCG**AACGGT**CGGGCCTGC-3' (SEQ ID NO:37)	remove *Acl*I	+
W358A	5'-GGCCTGCTGGCTTC**GC**GGAAAGCGACGACA-3' (SEQ ID NO:53)	none	-[c]
D362A	5'-GAAAGC**GACGCC**AATGACAAT-3' (SEQ ID NO:54)	*Bsa*HI	-
M366W	5'-ACGACAATGA**CAATTG**GGAAACAGCTTCGC-3' (SEQ ID NO:55)	*Mfe*I	++

[a] Underlined bases indicate the position of the introduced or eliminated restriction site. Base changes are in bold. [b] Indigo formation was monitored after 8 h as described in Materials and Methods. +++, colonies dark blue (corresponds to JM109(DE3)(pDTG141), expressing wild-type NDO; ++, colonies medium blue; +, colonies pale blue; -, no blue color (corresponds to negative control, JM109(DE3)(pT7-5). [c] Colonies pale blue after 12 h.

Regioisomers of biphenyl dihydrodiol were separated by PLC (1.0 or 2.0 mm thickness; E. Merck Industries, Inc., Gibbstown, N.J.) using multiple elution (3-4 developments) with chloroform-acetone (9:1). *cis*-Biphenyl 3,4-dihydrodiol was also purified by radial-dispersion chromatography (RDC) using a Chromatotron (Harrison Research, Palo Also, Calif.). Extracts in chloroform containing 0.1% triethylamine were applied to 2.0 mm-thick silica plates and eluted at a flow rate of 7 ml/min with a chloroform-acetone step gradient (0 to 15% acetone in 3% steps over 1 h; 0.1% triethylamine was present at each step). Fractions (8 ml) were analyzed by TLC, and those containing *cis*-biphenyl 3,4-dihydrodiol were combined and concentrated at 35 °C under reduced pressure.

Chiral stationary-phase HPLC was used to resolve the enantiomers of *cis*-naphthalene dihydrodiol. A Chiralcel OJ column (Chiral Technologies, Exton, Pa.) was used as described previously (S.M. Resnick, et al., *Appl. Environ. Microbiol.* **1994**, *60*, 3323-3328). Under these conditions, the (+)-(1*R*,2*S*)- and (-)-(1*S*,2*R*)- enantiomers of *cis*-naphthalene dihydrodiol eluted with retention times of 30 and 33 minutes, respectively. Proton (^{1}H) nuclear magnetic resonance (NMR) spectra were acquired on the Varian UNITY-500 500 MHz spectrometer in the College of Medicine NMR Facility at the University of Iowa. All spectra were obtained using an 8 second relaxation delay, a 5 second acquisition time, a spectral width of 12 ppm and a 90 degree pulse width of 6.6 microseconds. Samples were prepared as described by S.M. Resnick, et al., *Appl. Environ. Microbiol.* **1994**, *60*, 3323-3328). Optical rotations were determined at 25 °C using a Jasco P1020 polarimeter with a 589 nm Na lamp. The results are the average of rotations given by three independently purified *cis*-biphenyl 3,4-dihydrodiol samples. High resolution mass spectra were recorded (by Dr. Lynn Teesch, HR-MS facility, The University of Iowa) on a VG ZAB-HF mass spectrometer equipped with direct inlet probe. Absorbance spectra (200 to 350 nm) were recorded on a Beckman DU-70 spectrophotometer.

Chemicals. Naphthalene was obtained from Fisher Scientific Co., Pittsburg, Penn. Indole, biphenyl, phenanthrene, and 4-hydroxybiphenyl were purchased from Aldrich Chemical Co., Milwaukee, Wis. Synthetic (+/-)-*cis*-naphthalene dihydrodiol and homochiral (+)-*cis*-naphthalene dihydrodiol were

prepared as previously described (A.H. Jaffrey et al., *J. Org. Chem.*, **1974**, *39*, 1405-1407; S.M. Resnick and D.T. Gibson, *Biodegredations*, **1993**, *4*, 195-203). Synthetic *cis*-phenanthrene 9,10-dihydrodiol was provided by Dr. Derek Boyd.

Construction and preliminary analysis of modified NDO proteins. Nine positions near the active site in the α subunit of NDO were chosen for site-directed mutagenesis. Based on the crystal structure of NDO, Asn-201, Phe-202, Val-260, Trp-316, Phe-352, Trp-358 and Met-366 are located near enough to the mononuclear iron to interact with substrates in the active site. Asn-201 is positioned too far from the iron atom to be a ligand in the crystallized form of NDO, but was suggested as a possible ligand during some stage of the catalytic cycle. According to the NDO structure, Asp-362 is one of three amino acids that coordinate the iron at the active site. Asp-362 was replaced by alanine in order to disrupt iron coordination. Amino acid substitutions were also made at Thr-351, since the corresponding amino acid has be shown to be critical in determining polychlorinated biphenyl (PCB) congener specificity in biphenyl dioxygenase.

Site-directed mutations made in the α subunit of NDO are shown in **Table 5**. In most cases, small hydrophobic amino acids (alanine, valine, and leucine) were substituted for larger hydrophobic amino acids such as phenylalanine and tryptophan in order to change the size and/or shape of the active site pocket. In some cases, amino acid substitutions were chosen based on alignments of various related dioxygenase sequences.

Indigo formation was used as an initial screen for NDO activity. Freshly grown cells of JM109(DE3) carrying modified pDTG141 plasmids were incubated in the presence of indole. Most strains carrying mutant NDO enzymes formed blue colonies in the presence of indole. Strains producing NDO isoforms F202L and D362A formed white colonies, suggesting that either these enzymes were inactive or that indole was not a substrate for the modified enzymes. The strain carrying the W358A substitution in NDO formed pale blue colonies upon extended incubation with indole, indicating very weak activity with indole as substrate.

Production of mutant NDO α subunits. Formation of mutant α subunits was verified in Western blots using whole-cell protein samples from induced JM109(DE3) carrying modified pDTG141 plasmids. A monoclonal antibody specific for the α subunit of NDO was used R.E. Parales, et al., *J. Bacteriol.*, **1998**, *180*, 2337-2344). Results show that all mutant constructs formed full length α subunits and there were minor variations in the amounts of each mutant protein produced. More importantly, these show results demonstrate that the inability of isoforms D362A and F202L to produce products was not due to the absence of protein.

Biotransformations with naphthalene as substrate. Wild-type NDO converts naphthalene to *cis*-naphthalene 1,2-dihydrodiol. Biotransformations with naphthalene resulted in the formation of *cis*-naphthalene 1,2-dihydrodiol by all NDO isoforms with the exception of F202L and D362A, which formed no product. W358A transformations were very inefficient, with less than 5% of the substrate transformed within 15 h as judged by GC-MS analysis of extracted culture supernatants. The wild-type and all mutant NDO enzymes formed enantiomerically pure (>99%) (+)-(1*R*,2*S*)- *cis*-naphthalene dihydrodiol except for those with amino acid substitutions at Phe-352. Isoforms F352V and F352L formed 92% and 96% (+)-(1*R*,2*S*)- *cis*-naphthalene dihydrodiol, respectively. This result show the importance of a specific amino acid, Phe-352, in determining the enantioselectivity of NDO.

Biotransformations with biphenyl as substrate. Wild-type NDO oxidized biphenyl to two metabolites which were detected by TLC. The major metabolite (R_f, 0.2) and the minor metabolite (R_f, 0.18) dehydrated to phenolic products (M^+, 170) when analyzed by GC-MS. These results suggested that both metabolites were dihydrodiol isomers and this was confirmed by GC-MS of their stable respective phenyl boronic acid derivatives which gave molecular ions at m/e 274. The major metabolite (87% relative yield) had a retention time of 13.8 min and was identical to *cis*-2,3-dihydroxy-1-phenylcyclohexa-4,6-diene (*cis*-biphenyl 2,3-dihydrodiol) produced from biphenyl by *Sphingomonas yanoikuyae* B8/36 (formerly *Beijerinckia* sp. strain B8/36) (D.T. Gibson et al., *Biochem. Biophys. Res. Commun.* **1973**, *50*, 211-219). The minor product (13% relative

yield) had a retention time of 14.2 min and was identified as *cis*-3,4-dihydroxy-1-phenylcyclohexa-1,5-diene (*cis*-biphenyl 3,4-dihydrodiol; see below).

Isoforms F202L and D362A formed no products from biphenyl. Isoforms N201A and W358A formed only a trace amounts of *cis*-biphenyl 2,3-dihydrodiol. Amino acid substitutions at N201, F202, V260, W316, and T351 had slight effects on the regiospecificity of NDO as seen by the product distributions shown in **Figure 1**. However, both NDO isoforms with changes at position 352 formed *cis*-biphenyl 3,4-dihydrodiol as the major product. The isoform with the largest specificity change, F352V, formed 96% *cis*-biphenyl 3,4-dihydrodiol.

Identification and characterization of *cis*-biphenyl 3,4-dihydrodiol. The second product ran slightly slower than *cis*-biphenyl 2,3-dihydrodiol on TLC plates and, when analyzed by GC-MS as its PBA derivative, had a retention time of 14.2 min compared to 13.8 min for the PBA derivative of *cis*-biphenyl 2,3-dihydrodiol. The products formed from biphenyl by the F352V isoform were isolated by RDC. Approximately 140 mg of crude extract was applied to a 2.0 mm-thick silica chromatotron plate and eluted as described in the Materials and Methods to allow isolation of 40-60 mg pure *cis*-biphenyl 3,4-dihydrodiol and 1-2 mg of *cis*-biphenyl 2,3-dihydrodiol (fractions eluting before the 3,4-diol). The 3,4-regiochemistry of the diol and ^{1}H NMR shift assignments were established by chemical shift multiplicities and independent H-H decoupling experiments. The 3,4-regiochemistry of the dihydrodiol was apparent by decoupling at H-3 (4.31 ppm) which reduced the multiplicity of the H-2 signal (ddd, 6.16) to a singlet with fine splitting. Acid dehydration resulted in the formation of a product that coeluted with authentic 4-hydroxybiphenyl in TLC and GC-MS analyses.

Physical characteristics of the *cis*-biphenyl 3,4-dihydrodiol were as follows: λ_{max} [MeOH], 204, 228, and 276 nm, ϵ_{204} = 11,860, ϵ_{228} = 18,580, and ϵ_{276} = 4,336 $M^{-1}cm^{-1}$; calculated mass for the phenylboronate derivative, $^{12}C_{18}{}^{1}H_{15}{}^{16}O_2{}^{11}B$, 274.1160, found mass 274.1165; mass spectrum of phenyl boronate derivative m/z (relative intensity), 174 (M^+, 100), 170 (55), 152 (11), 142 (84), 115 (22), 77 (6); $[\alpha]_D$ –37.5 ± 4.4, n=3 (c 0.5, MeOH); ^{1}H NMR

(chloroform), δ 4.21 (ddd, J=6.4, 4.0, 1.5 Hz, H-4), 4.31 (dd, J=6.4, 4.2 Hz, H-3), 6.09 (ddd, J=9.8, 4.0, 0.8 Hz, H-5), 6.16 (ddd, J=4.2, 1.7, 0.7 Hz, H-2), 6.37 (dt, J=9.9, 1.6 Hz, H-6), 7.30 (tt, 1H aromatic-*p*), 7.37 (m, 2H, aromatic-*m*), 7.46 (m, 2H, aromatic-*o*).

Biotransformations with phenanthrene as substrate. Since many of the amino acid substitutions would be predicted to increase the size of the NDO active site, a larger substrate, phenanthrene, was tested. Identification of the three regioisomers of *cis*-phenanthrene dihydrodiol were carried out by comparing GC-MS data with *cis*-3,4-dihydroxy-3,4-dihydrophenanthrene (*cis*-phenanthrene 3,4-dihydrodiol) and *cis*-1,2-dihydroxy-1,2-dihydrophenanthrene (*cis*-phenanthrene 1,2-dihydrodiol) produced by *S. yanoikuyae* B8/36 (D.M. Jerina et al., *J. Am. Chem. Soc.* **1976**, *98*, 5988-5996), and synthetic *cis*-9,10-dihydroxy-9,10-dihydrophenanthrene (*cis*-phenanthrene 9,10-dihydrodiol). The PBA derivatives of *cis*-phenanthrene 9,10-dihydrodiol, *cis*-phenanthrene 3,4-dihydrodiol and *cis*-phenanthrene 1,2-dihydrodiol had GC retention times of 18.0, 19.1, and 20.2 min, respectively. Wild-type NDO from *Pseudomonas* sp. strain NCIB 9816-4 formed a 9:1 mixture of *cis*-phenanthrene 3,4-dihydrodiol and *cis*-phenanthrene 1,2-dihydrodiol (**Figure 2**). These results are similar to those obtained with NDO from *Pseudomonas* sp. strain 119 and biphenyl dioxygenase from *S. yanoikuyae* B8/36 (D.M. Jerina et al., *J. Am. Chem. Soc.* **1976**, *98*, 5988-5996).

With the exception of isoforms F202L and D362A, all mutant NDO enzymes formed products with phenanthrene as substrate. Amino acid substitutions at all positions changed product ratios to some extent. Isoforms V260A and W358A preferentially oxidized phenanthrene at the C3 and C4 positions, forming almost no *cis*-phenanthrene 1,2-dihydrodiol (**Figure 2**). Several isoforms, including N201A, N201S, F202V, W316A, T351R, F352V, F352L, and M366W produced a significantly greater proportion of *cis*-phenanthrene 1,2-dihydrodiol than did wild-type NDO (**Figure 2**). Of particular interest is the result with the F352V isoform. This enzyme had the opposite regioselectivity to wild-type NDO, forming 83% *cis*-phenanthrene 1,2-dihydrodiol in contrast to wild type, which formed 90% *cis*-phenanthrene 3,4-

dihydrodiol. The F352L isoform oxidized phenanthrene to a small amount (5% of the total product) of *cis*-phenanthrene 9,10-dihydrodiol (**Figure 2**).

Table 6 shows the amino acids in related dioxygenases that are located at positions corresponding to those mutated in NDO. Some amino acids listed in **Table 6** are conserved in all of the enzymes shown (Phe-202 and Asp-362). In other cases, amino acids are not conserved and an amino acid in NDO was changed to one present in one of the other enzymes. Some of the NDO mutations were chosen based on the identification of amino acids critical for determining substrate specificity in other dioxygenases. In NDO, Thr-351, when changed to Asn, had a minor effect on product formation from phenanthrene. Replacement of this amino acid with Arg in NDO had a slight effect on product formation from biphenyl and a larger effect when phenanthrene was provided as the substrate (**Figures 1 and 2**). This position corresponds to the important amino acid in $BPDO_{LB400}$ Asn-377 that was mentioned above.

Changes at Val-260 in NDO resulted in minor changes in product formation with biphenyl and phenanthrene. Substitution of Val for Asn260 in 2NTDO resulted in an enzyme that no longer oxidized the aromatic ring of 2-nitrotoluene, forming only the monooxygenation product 2-nitrobenzylalcohol (J.V. Parales and D.T. Gibson, *Abstracts of the 99th General Meeting of the American Society for Microbiology*, **1999**, Q-249, p-579). The opposite change in specificity did not occur with the NDO isoform V260N. Like wild-type NDO, the V260N isoform did not oxidize the aromatic ring of 2-nitrotoluene, but formed only 2-nitrobenzyl alcohol.

Toluene dioxygenase, which has a Trp residue at the position corresponding to 366 in NDO, dihydroxylates the aromatic ring of toluene to form *cis*-toluene dihydrodiol. However, the M366W isoform of NDO oxidized toluene to benzyl alcohol, the same product formed by the wild-type enzyme. Changing Trp-316 to Ala resulted in a minor change in regioselectivity with phenanthrene. Changing this conserved amino acid to Phe in 2NTDO had a slight effect on the stereochemistry of *cis*-naphthalene dihydrodiol formed from naphthalene (J.V. Parales and D.T. Gibson, *Abstracts of the 99th General Meeting of the American Society for Microbiology*, **1999**, Q-249, p-579).

TABLE 6. Comparison of amino acids at the active sites of selected dioxygenase α subunits

	Enzyme[a]						
Position[b]	NDO_{9816-4}	$2NTDO_{JS42}$	$DNTDO_{DNT}$	TDO_{F1}	$BPDO_{LB400}$	$BPDO_{KF707}$	NDO mutations
201	Asn	Asn	Asn	Gln	Gln	Gln	Ala, Gln, Ser
202	Phe	Phe	Phe	Phe	Phe	Phe	Leu, Val
260	Val	Asn	Val	Leu	Ser	Met	Ala, Leu, Asn
316	Trp	Trp	Phe	Trp	Trp	Trp	Ala
351	Thr	Ser	Ser	Thr	Asn	Thr	Asn, Arg, Ser
352	Phe	Ile	Thr	Phe	Phe	Phe	Leu, Val
358	Trp	Trp	Trp	Phe	Phe	Phe	Ala
362	Asp	Asp	Asp	Asp	Asp	Asp	Ala
366	Met	Met	Met	Trp	Trp	Trp	Trp

[a] Enzymes: 2NTDO, 2-nitrotoluene dioxygenase from *Pseudomonas* sp. strain JS42 (49); DNTDO, 2,4-dinitrotoluene dioxygenase from *Burkholderia* sp. strain DNT (64); TDO, toluene dioxygenase from *Pseudomonas putida* F1 (67); $BPDO_{LB400}$, biphenyl dioxygenase from *Burkholderia* sp. strain LB400 (18); $BPDO_{KF707}$, biphenyl dioxygenase from *Pseudomonas pseudoalcaligenes* KF707 (K. Taira et al., *J. Biol. Chem.*, 1992, *267*:4844-4853.)

[b] Position numbers refer to NDO. Alignments were carried out with the Pileup program (Wisconsin Sequence Analysis Package: Genetics Computer Group, Madison, Wisc.) using a gap weight of 3.5 and a gap length of 0.1.

The amino acid at position 352 appears to play an important role in controlling both the stereochemistry of *cis*-naphthalene dihydrodiol formed from naphthalene, as well as the regioselectivity with substrates such as biphenyl and phenanthrene. In addition, a product that is not made by wild-type NDO, *cis*-phenanthrene 9,10-dihydrodiol, was formed from phenanthrene by the F352L isoform.

To compare substrate specificities of NDO and the new NDO isoforms with those of the closely related enzymes 2NTDO and DNTDO, biotransformations were carried out with biphenyl and phenanthrene. Both were found to be poor substrates for 2NTDO and DNTDO. Both enzymes made a trace amount of *cis*-biphenyl 2,3-dihydrodiol from biphenyl, and DNTDO made a trace amount of phenanthrene 3,4-dihydrodiol from phenanthrene. It is not clear at this time why biphenyl and phenanthrene are such poor substrates for 2NTDO and DNTDO.

In NDO, Asp-205 is located between the two redox centers at the junction of two adjacent α subunits. Substitution of Asp-205 by glutamine resulted in an isoform of NDO with no activity (R.E. Parales, et al. *J. Bacteriol.*, **1999**, *181*, 1831-1837). In the glutamine-containing enzyme, electron transfer between the Rieske center and the mononuclear iron was shown to be blocked, indicating that Asp-205 is essential for this electron transfer step to occur (R.E. Parales, et al. *J. Bacteriol.*, **1999**, *181*, 1831-1837). Iron at the active site of NDO is coordinated by His-208, His-213, and Asp-362. All three of these residues are conserved in the ring-hydroxylating dioxygenases whose sequences have been determined to date.

The corresponding histidine residues in toluene dioxygenase, from *P. putida* F1 (His-222 and His 228) were replaced with alanine residues and these substitutions resulted in completely inactive enzymes (H. Jiang, et al., *J. Bacteriol.* **1996**, *178*, 3133-3139). The inability to detect products from four different substrates indicates that substitution of Ala at position 362 results in an inactive form of NDO. No activity was detected in crude cell extracts of the D362A isoform with either oxygen uptake assays or product formation assays

with ^{14}C-naphthalene. These results are consistent with the identification of Asp-362 as a ligand to the mononuclear iron at the active site.

Asn-201, a possible fourth iron-coordinating amino acid was observed in the crystal structure of NDO. This residue was too far from the iron atom to serve as a ligand in the crystallized form of the enzyme, but was suggested as a possible ligand during a step in the catalytic cycle (B. Kauppi, et al., *Structure*, **1998**, *6*, 571-586). Amino acid substitutions at Asn-201 resulted in enzymes with reduced but significant activity, indicating that this residue does not participate in the coordination of iron at the active site. Crude cell extracts of the N201A and N201Q isoforms had 5-10% of the activity of wild-type NDO. Results presented in **Figures 1 and 2** suggest that Asn-201 may play a minor role in determining regioselectivity with biphenyl and phenanthrene as substrates. However, Asn-201 may be more important for maintaining appropriate interactions between α subunits through its hydrogen bond with Tyr-103 near the Rieske center in an adjacent α subunit (B. Kauppi, et al., *Structure*, **1998**, *6*, 571-586). Substitution of an alanine at Asn-201 would disrupt this hydrogen bond and could affect the flow of electrons from the Rieske center to the mononuclear iron, thus reducing enzyme activity. The incorporation of the larger Gln residue at this position may prevent the normal interaction of α subunits even though Gln would be capable of forming a hydrogen bond with Tyr-103. The N201S isoform showed 35-40% of the wild-type NDO activity, indicating that serine is a reasonably good substitute for Asn at this position (B.V. Plapp, *Methods Enzymol.*, **1995**, *249*, 91-119).

Of the three substrates tested, the most significant effects of mutations at the active site were observed with the largest substrate, phenanthrene. This is not an unexpected result since the substrate pocket is of limited size and larger substrates are likely to come in contact with more amino acids in the active site. Many of the mutations involved the substitution of a small hydrophobic amino acid for a larger one, and in most cases this type of substitution did not severely reduce the activity of the enzyme as can sometimes occur (M.S. Caffrey, *Biochimie*, **1994**, *76*, 622-630). However, one exception was the substitution of Trp-358 by Ala, which resulted in an enzyme with very

poor activity with naphthalene and biphenyl, but somewhat better activity with the larger substrate phenanthrene.

Another exception was isoform F202L, which failed to form products with all substrates tested. The reason that this substitution resulted in an inactive enzyme while isoform F202V had good activity is not understood. Somewhat surprisingly, mutations that introduced changes in polarity or charge (V260N, T351R) resulted in enzymes with good activity toward hydrophobic substrates. In general, most changes at the active site, with the exception of those that affect iron binding (Asp-362) and electron transfer (Asp-205) were tolerated well, suggesting that there is significant flexibility in the range of amino acids that can be introduced at the active site. This suggests that oxygenases with novel catalytic capabilities can be generated by introducing single or multiple mutations near the active site.

EXAMPLE 7 **Other Mutants Related To SEQ ID NO:25**

Using procedures similar to those described in Example 1 (as described below), NDO mutant genes encoding glycine, alanine, threonine, leucine, isoleucine, tryptophane, or tyrosine instead of phenylalanine at position 352 were also prepared. Table 7 shows the SEQ ID No's for these DNA sequences and for proteins they encode.

Table 7

Amino Acid At Position 352	SEQ ID NO: For Modified Sequence	SEQ ID NO: For Corresponding Polypeptide
glycine	SEQ ID NO:27	SEQ ID NO:32
alanine	SEQ ID NO:28	SEQ ID NO:33
threonine	SEQ ID NO:29	SEQ ID NO:34
leucine	SEQ ID NO:30	SEQ ID NO:35
isoleucine	SEQ ID NO:31	SEQ ID NO:36
tyrptophane	SEQ ID NO:56	SEQ ID NO:58
tyrosine	SEQ ID NO:57	SEQ ID NO:59

Activity of modified NDO proteins. The formation of indigo from indole was used to screen for NDO activity. Freshly grown colonies JM109(DE3) carrying modified pDTG141 plasmids were incubated in the presence of indole. Strains producing NDO enzymes with the mutations F352W and F352Y formed white colonies, suggesting that these enzymes were inactive or that indole was no longer a substrate for the modified enzymes. All other NDO isoforms constructed in this study appeared to be active. **Table 8** shows the substitutions in the α subunit of NDO generated by site-directed mutagenesis.

Table 8

Mutation	Mutagenic Oligonucleotide	SEQ ID NO:	Indigo Formation
F352G	5'-GTTCAGCGAACGGGCGGGCCTGCTGG-3'	60	+
F352A	5'-GTTCAGCGAACGGCCGGGCCTGCTGG-3'	61	+
F352T	5'-GTTCAGCGAACGACCGGGCCTGCTGG-3'	62	+
F352I	5'-GTTCAGCGAACGATCGGGCCTGCTGG-3'	63	+
F352L	5'-TTCAGCGAACGCTCGGGCCTGC-3'	52	+
F352W	5'-GTTCAGCGAACGTGGGGGCCTGCTGG-3'	64	-
F352Y	5'-TTCAGCGAACGTACGGGCCTGCTGG-3'	65	-

Underlined bases in Table 8 indicate the position of the eliminated restriction site, *Acl*I. BAase changes are in bold. Indigo formation was monitored after 8 hours; (+) indicates pale blue colonies, (-) indicates no blue color.

Regioselectivity of modified NDO proteins. Biotransformations with naphthalene resulted in the formation of *cis*-1,2-dihydroxy-1,2-dihydronaphthalene (*cis*-naphthalene dihydrodiol) by all NDO isoforms with substitutions at position 352 except F352Y, which formed no product. NDO-F352W transformations were very inefficient. In contrast to wild-type NDO, all enzymes with amino acid substitutions at position 352 formed small amounts of the (-)-enantiomer of *cis*-naphthalene dihydrodiol from naphthalene as determined from chiral HPLC analysis **(Table 9)**.

Table 9

NDO Enzyme	cis-Naphthalene 1,2-dihydrodiol	cis-Biphenyl 2,3-dihydrodiol	cis-Biphenyl 3,4-dihydrodiol
NDO (wild type)	>99% (+)-(1R,2S)	>95% (+)-(2R,3S)	>98% (+)-(3R,4S)
F352G	98% (+)-(1R,2S)	>95% (+)-(2R,3S)	60% (+)-(3R,4S)
F352A	96% (+)-(1R,2S)	>95% (+)-(2R,3S)	65% (+)-(3R,4S)
F352T	93% (+)-(1R,2S)	>95% (+)-(2R,3S)	60% (-)-(3S,4R)
F352V	92% (+)-(1R,2S)		77% (-)-(3S,4R)
F352I	94% (+)-(1R,2S)	>95% (+)-(2R,3S)	53% (+)-(3R,4S)
F352L	96% (+)-(1R,2S)	>95% (+)-(2R,3S)	70% (+)-(3R,4S)

When biphenyl was used as a biotransformation substrate, wild-type NDO converted it to an 87:13 mixture of *cis*-biphenyl 2,3-dihydrodiol and *cis*-biphenyl 3,4-dihydrodiol. However, a major change in regioselectivity with biphenyl was seen when amino acid substitutions were introduced at F352. All active mutant NDO enzymes with changes at this position formed *cis*-biphenyl 3,4-dihydrodiol as the major product **(Table 10)**. The F352Y isoform formed no detectable product from biphenyl, and F352W formed only a trace amount of *cis*-biphenyl 2,3-dihydrodiol.

Table 10

Enzyme	Biphenyl 2,3-diol	Biphenyl 3,4-diol	phenanthrene 3,4-diol	phenanthrene 1,2-diol	phenanthrene 9,10-diol
NDO	87	13	90	10	-
F352G	32	68	79	21	-
F352A	23	77	53	47	-
F352T	8	92	59	41	-
F352V	4	96	17	83	-
F352L	15	85	64	31	5
F352I	17	83	76	24	-

Like wild-type NDO, isoforms F352G, F352A, F352T, F352I and F352L formed *cis*-phenanthrene 3,4-dihydrodiol as the major product from phenanthrene, although product ratios varied significantly depending on the enzyme (**Table 10**). The F352V isoform had the opposite regioselectivity, forming primarily (83%) *cis*-phenanthrene 1,2-dihydrodiol. Isoforms F352W and F352Y did not form detectable amounts of product from phenanthrene.

The enantiomeric composition of *cis*-biphenyl 2,3-dihydrodiol was unaffected by amino acid substitutions at this position, but that of the *cis*-biphenyl 3,4-dihydrodiol was significantly different in all cases from that formed by wild type (**Table 9**). Isoforms F352V and F352T formed the opposite enantiomer of *cis*-biphenyl 3,4-dihydrodiol as wild-type NDO (**Table 9**).

Absolute stereochemistry of *cis*-biphenyl 3,4-dihydrodiol. The formation of diastereomeric 2-(1-methoxyethyl)-phenyl boronic acid (MPBA) derivatives of the F352V-generated *cis*-biphenyl 3,4-dihydrodiol provided a means for determining the enantiomeric purity of the compound. The results also allow an empirical prediction of absolute configuration based on trends for vicinal *cis*-diols with a benzylic hydroxymethine (S.M. Resnick, et al., *J. Org. Chem.*, **1995**, *60*, 3546-3549). These trends were employed in the absence of MPBA directional shift data for a series of *cis*-3,4-dihydrodiols of known absolute configuration. Proton (^{1}H) nuclear magnetic resonance (NMR) spectra were acquired on the Bruker AMX-600 MHz NMR spectrometer at 600.14 MHz

in the University of Iowa High-Field NMR Facility. All spectra were obtained using a 14 s recovery delay, a 4.06 s acquisition time, a spectral width of 13.4 ppm and a 90 degree pulse width of 7.5 μs. Samples were prepared as previously described (S. M. Resnick et al., *J. Org. Chem.*, **1995**, *60*, 3546-3549; S.M. Resnick et al., *Appl. Enviorn. Microbiol.*, **1994**, *60*, 3323-3328). ^{1}H-NMR analysis (d_6-benzene) showed that the methoxy signal of the derivative formed with the (-)-*cis*-3,4-biphenyl dihydrodiol and (*S*)-MPBA was shifted downfield ($\Delta\delta$ +21 ppb; 3.1987 ppm) relative to the corresponding signal of the (*R*)-MPBA derivative. The enantiomeric purity of the major diol was approximately 75%, based on integration of the methoxy groups of the major and minor MPBA diastereomers. This result confirms the data obtained by chiral HPLC analysis (**Table 9**). The downfield shifted methoxy signal for the (*S*)-MPBA derivative would indicate an *S*-configuration at the benzylic carbon for a 2,3-dihydrodiol. Application of this trend to the hydroxymethine nearest the benzylic position allows the suggestion of *S*-stereochemistry at C-3, and an absolute configuration of (-)-*cis*-(3*S*,4*R*)-biphenyl dihydrodiol.

Preparation of enantiomerically pure (-)-*cis*-(3*S*,4*R*)-biphenyl dihydrodiol. Three dihydrodiol dehydrogenases were tested for the ability to oxidize the (+)- and (-)-enantiomers of *cis*-biphenyl 2,3- and 3,4- dihydrodiols. Toluene dihydrodiol dehydrogenase from *Pseudomonas putida* F1 (J.E. Rogers and D.T. Gibson, *J. Bacteriol.*, **1977**, *130*, 1117-1124) was shown to specifically attack the (+)-enantiomers of both dihydrodiols at a significant rate. Plasmid pDTG511 carries the *todD* gene encoding toluene dihydrodiol dehydrogenase from *P. putida* F1 (G.J. Zylstra and D.T. Gibson Arromatic Hydrocarbon Degredation: a molecular approach, p. 183-203, in K. Setlow (ed.) Genetic Enginering: Principles and Methods, vol. 13, Plenum Press, New York). A bacterial strain was constructed which produces the F352V isoform of NDO and toluene dihydrodiol dehydrogenase. When this strain, JM109(DE3)(pDTG141-F352V)(pDTG511), was used in biotransformations with biphenyl, the only dihydrodiol detected by chiral HPLC anaylsis was (-)-*cis*-biphenyl 3,4-dihydrodiol. The (+)-enantiomers of *cis*-biphenyl 2,3-dihydrodiol and *cis*-biphenyl 3,4-dihydrodiol were completely converted to the respective catechols,

compounds which were easily separated from the dihydrodiol by preparative thin layer chromatography.

Absolute stereochemistry of *cis*-phenanthrene 3,4-dihydrodiol and 1,2-dihydrodiol. The formation of diastereomeric MPBA derivatives of the F352V-generated mixture of *cis*-phenanthrene dihydrodiols provided a means for determining the enantiomeric purity of the compounds. The results also allow an empirical prediction of absolute configuration based on trends for vicinal *cis*-diols with a benzylic hydroxymethine (S. M. Resnick et al., *J. Org. Chem.*, **1995**, *60*, 3546-3549). Samples were prepared as previously described (S. M. Resnick et al., *J. Org. Chem.*, **1995**, *60*, 3546-3549; S.M. Resnick et al., *Appl. Enviorn. Microbiol.*, **1994**, *60*, 3323-3328). Proton (^{1}H) nuclear magnetic resonance (NMR) spectra were acquired as described above. An upfield directional shift for the methoxy signal of many (*S*)-MPBA deriviatives of polyaromatic *cis*-1,2-dihydrodiols is indicative of an *R*-configuration at the benzylic position of the *cis*-diol. For the *cis*-phenanthrene 1,2-dihydrodiol, the methoxy signal of the (*S*)-MPBA deriviative was +72 ppb downfield from the corresponding signal of the opposite diastereomer formed with (*R*)-MPBA and predicting an *S*-configuration at the benzylic center.

Based on trends previously documented, the absolute configuration of the major dihydrodiol formed by F352V from phenanthrene is *cis*-(1*S*,2*R*)-phenanthrene dihydrodiol (91% e.e., approx. 83% relative yield). The facial selectivity in this case was the opposite to that shown for wild type biphenyl dioxygenase from *Sphingomonas yanoikuyae* B8/36. Analysis of the (±)-MPBA derivative of the isolated *cis*-phenanthrene dihydrodiol fraction formed by B8/36 showed resolution of the mixed racemates (of 3,4- and 1,2-diols) with minor methoxy signals of the 1,2-diol at 3.148 and 3.220 ppm. The same sample derivatized with (*S*)-MPBA showed the upfield shift for the methoxy signal at 3.148 ppm which corresponds to the an *R*-configuration of the benzylic center consistent with and confirmed by the previously determined (1*R*,2*S*)-configuration (M. Koreeda et al., *J. Org. Chem.*, **1978**, *43*, 1023-1027). The results of the above stereochemical correlation also suggest that the empirical application of the trends in the directional shifts of polycyclic aromatic diols

appears to be valid for both the "bay-region" *cis*-3,4- and "non-bay region" *cis*-1,2- dihydrodiols of phenanthrene.

The minor diol formed from phenanthrene by F352V was identified as *cis*-(3*S*,4*R*)-phenanthrene dihydrodiol (>95% e.e., 17% relative yield). This assignment is based on the correlation of the methoxy signal at 3.115 ppm (but not 3.241 ppm) in the (*S*)-MPBA derivative of the F352V minor phenanthrene 3,4-dihydrodiol with that of the identical directional shifts of the known B8/36 *cis*-phenanthrene 3,4-dihydrodiol derivatives.

Relative activities of the mutant NDO enzymes. Cultures (50 ml in 500 ml flasks) were grown and induced, and biotransformations with naphthalene or biphenyl were initiated as described previously. Samples (1 ml each) were taken at 30 minute intervals over a period of 5 hours. Cells were removed by centrifugation and pellets were stored at –20°C for protein determinations. *cis*-Naphthalene dihydrodiol formation was monitored at 262 nm (ϵ_= 8114 $M^{-1}cm^{-1}$). *cis*-Biphenyl 2,3-dihydrodiol formation was monitored at 303 nm (ϵ_= 13,600 $M^{-1}cm^{-1}$; (2)). *cis*-Biphenyl 3,4-dihydrodiol formation was monitored at 276 nm (ϵ_= 4340 $M^{-1}cm^{-1}$; (5)) using a correction for the absorbance of *cis*-biphenyl 2,3-dihydrodiol at this wavelength. The extinction coefficient of *cis*-biphenyl 2,3-dihydrodiol at 276 nm (the λmax of of *cis*-biphenyl 3,4-dihydrodiol) was determined to be 7950 $M^{-1}cm^{-1}$ using purified *cis*-biphenyl 2,3-dihydrodiol from *S. yanoikuyae* B8/36 (2). The concentration of *cis*-biphenyl 3,4-dihydrodiol was calculated using the ratios of products formed by each mutant enzyme and subtracting the contribution of *cis*-biphenyl 2,3-dihydrodiol. Absorbance of the negative control strain (JM109(DE3)(pT7-5) was subtracted at each time point. Protein concentrations were determined by the method of Bradford (M.M. Bradford, *Anal. Biochem.* **1976**, *72*, 248-254) after boiling cell pellets for 1 h in 0.1 N NaOH. Bovine serum albumin was used as the standard. Rates reported are the averages of three independent experiments.

The in vivo rates of formation of *cis*-naphthalene dihydrodiol by wild-type and mutant NDO enzymes are shown in **Table 11**.

Table 11

NDO Enzyme	Naphthalene dihydrodiol formation: Specific Activity (nmol/min/mg)	Naphthalene dihydrodiol formation: Relative Activity (%)	Biphenyl 2,3-dihydrodiol formmation: Specific Activity (nmol/min/mg)	Biphenyl 2,3-dihydrodiol formmation: Relative Activity (%)	Biphenyl 3,4dihydrodiol formation: Specific Activity (nmol/min/mg)	Biphenyl 3,4dihydrodiol formation: Relative Activity (%)
NDO	20.4	100	4.81	100	0.95	100
F352G	7.4	37	<0.05	<1	<0.25	<26
F352A	9.9	49	0.07	1	0.25	26
F352T	15.6	77	0.14	3	0.73	77
F352V	16.9	83	0.15	3	0.94	99
F352I	16.5	81	0.16	3	0.74	78
F352L	19.5	96	0.21	4	0.82	86

The F352L isoform produced *cis*-naphthalene dihydrodiol at wild-type rates, while the F352T, F352V, and F352I isoforms were slightly less efficient, with rates 75-85% that of wild-type NDO. The F352G and F352A enzymes were the least efficient in catalyzing this reaction. A similar trend is seen in the rates of formation of *cis*-biphenyl 3,4-dihydrodiol from biphenyl by the enzymes with substitutions at postion 352 (**Table 11**). The F352T, V, I and L isoforms formed product at slightly reduced rates compared to wild-type NDO, while F352A was significantly slower and F352G rates were not measurable. In contrast, all enzymes with substitutions at position 352 were severely defective in forming *cis*-biphenyl 2,3-dihydrodiol from biphenyl (**Table 11**). These studies demonstrate that the amino acid substitutions at position 352 result in enzymes with a decreased tendency to oxidize at the 2,3- position of biphenyl. However, the rate of oxidation at the 3,4- position of biphenyl was not improved in the mutant enzymes.

All publications, patents, and patent documents are incorporated by reference herein, as though individually incorporated by reference. The invention has been described with reference to various specific and preferred embodiments and techniques. However, it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention.